AFRL-HE-AZ-TP-2007-0002

The Effectiveness of a PC-Bases C-130 Crew Resource Management Aircrew Training Device

Robert T. Nullmeyer
Air Force Research Laboratory
6030 S. Kent Street
Mesa, AZ 85212

V. Alan Spiker
Anacapa Sciences, Inc.
301 E. Carrillo St.
Santa Barbara, CA 93101

Katharine C. Golas &
Ryan C. Logan
Southwest Research Institute
6220 Culebra Road
San Antonio, TX 78228

Larry Clemons
Air Education and Training Command
Randolph AFB, TX 78150

December 2006
Final Report for March 2005 to December 2006

Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Readiness Research Division

NOTICES

AFRL-HE-AZ-TP-2007-0002 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//signed//
ROBERT NULLMEYER
Lab Contract Monitor

//signed//
HERBERT H. BELL
Technical Advisor

//signed//
DANIEL R. WALKER, Colonel, USAF
Chief, Warfighter Readiness Research Division
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
6-12-2006	Conference Proceeings	March 2005 to Dec 2006

4. TITLE AND SUBTITLE
The Effectiveness of a PC-Bases C-130 Crew Resource Management Aircrew Training Device

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
62202F

6. AUTHOR(S)
Robert T. Nullmeyer
V. Alan Spiker
Katherine C. Golas
Ryan C. Logan
Larry Clemons

5d. PROJECT NUMBER
1123

5e. TASK NUMBER

5f. WORK UNIT NUMBER
1123AE08

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Training Research Division
6030 South Kent Street
Mesa AZ 85212-6061

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Training Research Division
6030 South Kent Street
Mesa AZ 85212-6061

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL; AFRL/HEA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-HE-AZ-TP-2007 - 0002

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Inadequate crew resource management (CRM) behaviors are still cited as causal factors in most military and commercial aircraft mishaps despite mandatory CRM training in virtually all aviator training programs, suggesting a need to explore alternative approaches. A low-cost, PC-based simulator was designed to elicit the communication and crew coordination behaviors associated with instrument and visual airdrop missions. These targeted behaviors were frequently addressed in instructor comments from earlier C-130 student training records, especially for navigators and copilots. The effectiveness of instruction using this device was evaluated. Treatment group students received a four hour training profile before their first airdrop flight while control group students did not. Multiple measures of effectiveness were tracked. Instructors and students rated training effectiveness using 5-point Likert scales. Ratings from both groups were significantly greater than "3" (neutral) for task management, communication, and crew coordination. In addition, instructors reported that the experience was a good use of instructor and student time. Detailed CRM proficiency data were collected during the first subsequent airdrop flight. Positive transfer of training was substantiated by a multivariate analysis of variance. CRM performance ratings during this flight were significantly higher for treatment group students than their for control group peers. Higher performance grades in training records were also observed for treatment group students in all CRM skill areas through subsequent flights, with fewer sorties to criterion for communication, crew coordination, task management, and decision making for both navigators and copilots.

Empirical CRM training effectiveness data are rare. This paper addresses the effectiveness of instruction using a PC-based simulator to develop teamwork skills and provides a template for measuring "soft skills" in operational environments using a combination of focused, study-specific data collection instruments and existing student training records. Each provided unique insights regarding benefits and limitations of PC-based CRM training.

15. SUBJECT TERMS
Crew resource management; PC-based simulator; training effectiveness evaluation; C-130 training; team trainin

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Robert Nullmeyer
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED	20	**19b. TELEPHONE NUMBER** *(include area code)* 480-988-6561 x283

This page intentionally left blank.

The Effectiveness of a PC-Based C-130 Crew Resource Management Aircrew Training Device

Robert T. Nullmeyer
Air Force Research Laboratory
6030 S. Kent Street
Mesa, AZ 85281

V. Alan Spiker
Anacapa Sciences, Inc.
301 E. Carrillo St.
Santa Barbara, CA 93101

Katharine C. Golas &
Ryan C. Logan
Southwest Research Institute
San Antonio TX

Larry Clemons
Air Education and Training Command
Randolph AFB TX

ABSTRACT

Inadequate crew resource management (CRM) behaviors are still cited as causal factors in most military and commercial aircraft mishaps despite mandatory CRM training in virtually all aviator training programs, suggesting a need to explore alternative approaches. A low-cost, PC-based simulator was designed to elicit the communication and crew coordination behaviors associated with instrument and visual airdrop missions. These targeted behaviors had frequently been cited as problematic by instructors in earlier C-130 student training records, especially for navigators and copilots. The effectiveness of instruction using this device was evaluated. Treatment group students received a four hour training profile before their first airdrop flight while control group students did not. Multiple measures of effectiveness were tracked. Instructors and students rated training effectiveness using 5-point Likert scales. Ratings from both groups were significantly greater than "3" (neutral) for task management, communication, and crew coordination. In addition, instructors reported that the experience was a good use of instructor and student time. Detailed CRM proficiency data were collected during the first subsequent airdrop flight. Positive transfer of training was substantiated by a multivariate analysis of variance. CRM performance ratings during this flight were significantly higher for treatment group students than their for control group peers. Higher performance grades in training records were also observed for treatment group students in all CRM skill areas through subsequent flights, with fewer sorties to criterion for communication, crew coordination, task management, and decision making for both navigators and copilots.

Empirical CRM training effectiveness data are rare. This paper addresses the effectiveness of instruction using a PC-based simulator to develop teamwork skills and provides a template for measuring "soft skills" in operational environments using a combination of focused, study-specific data collection instruments and existing student training records. Each provided unique insights regarding benefits and limitations of PC-based CRM training.

ABOUT THE AUTHORS

Robert Nullmeyer is a senior research psychologist with the Air Force Research Laboratory at Mesa, AZ, where he has been since 1981. He conducts research on training effectiveness and crew resource management. He earned a Ph.D. in Experimental Psychology from the University of New Mexico in 1979.

V. Alan Spiker is a principal scientist at Anacapa Sciences, Inc., Santa Barbara, CA. He is responsible for human factors research associated with advanced technology and has conducted field studies on mission planning and command and control systems for all branches of service. He has a Ph.D. in Experimental Psychology from the University of New Mexico and has been with Anacapa Sciences since 1982.

Katherine Golas is Vice President of the Training, Simulation and Performance Improvement Division at Southwest Research Institute, where she has worked since 1988. She directs a staff of approximately 100 scientists, engineers, and instructional technologists in the design and development of training and simulator programs. She has a Ph.D. in Instructional Systems from Florida State University.

Ryan C. Logan is a project manager in the Training, simulation, and Performance Improvement Division at Southwest Research Institute. He has five years of experience in modeling and development of simulation software. He received a Bachelor of Environmental Design and M.S. in Visualization Sciencesfrom Texas A&M University.

The Effectiveness of a PC-Based C-130 Crew Resource Management Aircrew Training Device

Robert T. Nullmeyer
Air Force Research Laboratory
6030 S. Kent Street
Mesa, AZ 85281

V. Alan Spiker
Anacapa Sciences, Inc.
301 E. Carrillo St.
Santa Barbara, CA 93101

Katharine C. Golas & Ryan C. Logan
Southwest Research Institute
San Antonio TX

Larry Clemons
Air Education and Training Command
Randolph AFB TX

INTRODUCTION

The central role of human error in flight mishaps is well documented. Helmreich and Fouchee (1993) reported that flight crew actions were causal in more than 70% of worldwide air carrier accidents from 1959 to 1989 involving aircraft damage beyond economic repair. More recently, Montimy (2005) reported major or causal human factors in the majority of Air Force Class A mishaps (destroyed aircraft, $1 million damage or fatality) and over 90% of fatal mishaps from 1991 to 2003 with little change over that time period, and White (2006) reported steadily increasing Crew Resource Management (CRM)-related Class A mishap rates over the past decade in the U.S. Navy.

Hendy, Thompson, Fraser, Jamieson, Comeau, Mack, Paul, and Brooks (1998) reviewed seven Canadian Air Force CC-130 mishaps from 1980-1993. Human factors issues were implicated in each. The majority of mishaps involved a breakdown of crew coordination, and decision making was such a central problem area that the study team recommended replacing the aircrew coordination training (ACT) program with decision-centric human factors training (Hendy & Ho, 1998).

In a review of Air Force C-130 Class A mishaps from 1992-2002, Nullmeyer, Stella, Flournoy, and White (2003) found very similar patterns. Causal CRM problems were common. Risk assessment or decision making behaviors alone were causal or major factors in 93% of these mishaps.

Some researchers have argued that there is no evidence to support CRM training effectiveness (Komich, 1997; Simmon, 1997). Based on a recent, comprehensive review of the CRM literature, Salas, Wilson, Burke, Whiteman and Howse (in press) concluded that, while CRM training is effective at some levels, "the picture is not as clear as it should be after 20 years" regarding transfer to the operational environment. In fact, transfer of CRM training to the aircraft is rarely addressed.

The 314th Airlift Wing (314 AW) submitted a proposal to the Education and Training Technology Application Program (ETTAP) to develop and evaluate a low-cost, PC-based aircrew training device (PC-ATD) to reinforce the CRM skills and tactical knowledge required in the tactical flying phase of the C-130 mission qualification course. They reported that students often find the transition from ground-based instruction to performance in the aircraft to be a rather broad leap and further reported that C-130 qualification and mission qualification students frequently require additional training in teamwork skills.

The proposed PC-ATD would enable pilot, co-pilot, navigator, and flight engineer students to practice CRM, visual airdrop, visual low-level map-reading skills, time over target (TOT) control, Station Keeping Equipment (SKE) procedures, checklists, radio procedures, and navigation systems management. ATD instruction would occur immediately prior to the first inflight airdrop. This added practice was expected to significantly enhance aircrew readiness by providing a training bridge between classroom/full mission simulator training and the aircraft immediately prior to the first airdrop flight.

Using the high-fidelity Weapon System Trainer (WST) to accomplish this "top off" instruction was not feasible due to high student loads and the collective training requirements levied by 6 flying C-130 squadrons. The PC-ATD was seen as a low-cost alternative to enhance CRM skills, especially aircrew situation awareness and coordinated crew responses, through practice in a realistic simulation of IFR and VFR air drop missions. Savings were anticipated through reduced "additional training" flights required

by students who would needed more practice to reach required proficiency levels.

An analysis of written flight instructor comments regarding student performance (Nullmeyer & Spiker, 2002) corroborated the main points made in the proposal. Instructor comments have several interesting properties. First, they are generated by experts in the students' crew positions. Second, they require added effort, and as a result, it is likely that they reflect observations regarding the student that instructors view are relatively important. Instructors in most crew positions provide fairly rich narratives of student strengths and weaknesses in these comments. The majority of comments (about 60%) are positive, but both positive and negative comments are frequently recorded.

Frequencies of negative comments in C-130 student records over a four month period in FY 2003 are summarized in Table 1. These frequencies are grouped by student training courses (columns) and by traditional CRM skill areas (rows in bold print). Five categories of students were included in this summary:

Aircraft Commander Qualification (ACQ) students – 28 experienced C-130 copilots upgrading to aircraft commander
Copilot Initial Qualification (CIQ) students – 27 recent undergraduate pilot training graduates qualifying as C-130 copilots
Navigator (NAV) students – 18 undergraduate navigation training graduates qualifying as C-130 navigators
Flight Engineer (FE) students – 25 enlisted personnel with technical backgrounds who were new to flying and qualifying as C-130 flight engineers
Loadmaster (LM) students – 28 enlisted personnel qualifying as C-130 loadmasters

The first six rows in Table 1 reflect the CRM areas in the Air Force Instruction (AFI) that defines CRM training (AFI 11-290) and the numbers of negative instructor comments directly related to each area. Frequently cited non-CRM instructor comment areas are listed in brackets. About half of the negative instructor comments in the time period reviewed reflected shortfalls in classic CRM skills. This is consistent with the need statement provided by the 314th AW that CRM skills represented areas where student behaviors could be improved. Among CRM skill areas, situation awareness (SA) was frequently problematic for student pilots upgrading to aircraft commander. Crew coordination appeared to be particularly challenging for student copilots and navigators. Task management comments for FEs were almost all related to checklists. Instructor comments of any sort (positive or negative) were rare in loadmaster student records.

Negative instructor comments were also common in several non-CRM areas. Aircraft handling and "stick and rudder" problems were commonly cited, especially for students who were upgrading to aircraft commander. Entering planning data was frequently problematic for FE students.

METHODS

The Study Plan

The formal study plan (Leonard, 2005) stated the project objective is to provide a proof-of-concept test to investigate the ability of a PC-based training device to improve teamwork skills for C-130 mission qualification students in the flightline phase of training. The core hypothesis was that instruction in the PC-ATD would enhance training by reinforcing cognitive skills (CRM, aircrew situational awareness, formation procedures, threat recognition/reaction, and coordinated crew responses) through practice in a realistic simulation of the airlift mission environment. The largest gains in student performance were expected in copilot and navigator crew positions given the unusually high numbers of negative comments for these crew positions in these areas. Based on the specific problems cited by instructors in earlier student records, the targeted skills in this proof-of-concept study involved interactions among student crewmembers – assertiveness, crosstalk, backing up other crewmembers, and communication discipline.

Table 1. Frequencies of Negative Instructor Comments for CRM in C-130 Student Records

Skill Area	ACQ (n=28)	CIQ (n=27)	NAV (n=18)	FE (n=25)	LM (n=28)
Task Management	**31**	**24**	**23**	**48**	**15**
Situation Awareness (SA)	**114**	**36**	**39**	**3**	**3**
Communication	**5**	**32**	**47**	**10**	**3**
Crew Coordination	**35**	**72**	**107**	**12**	**4**
Decision Making/Risk Assessment	**4**	**3**	**4**	**1**	**1**
Mission Planning	**26**	**7**	**17**		**3**
[Aircraft Handling]	*[138]*	*[33]*			

[Stick/Rudder/Throttle Input]	*[61]*	*[22]*		
[Planning Data Entry]		*[1]*	*[2]*	*[46]*

The study plan called for a classic treatment group/control group design. In-flight CRM proficiency of crews who had practiced airdrop missions in the PC-ATD would be compared with the CRM proficiency exhibited in-flight by a second set of student crews who did not receive this practice. Detailed observations of CRM skills would be documented by instructors during the first airdrop flight in the C-130 aircraft for both test group and control group students. The plan specified four measures of effectiveness:

- **Student feedback** questionnaires were administered after the student's first airdrop in the actual C-130 aircraft and addressed the value of the PC-ATD experience and the fidelity of the device.

- **Instructor assessments** regarding training value, training impacts, and implementation strategies were solicited from those who had used the device to instruct treatment group students. Assessments were gathered once after training ended for the treatment group.

- **Instructor ratings of student CRM skills in-flight** were viewed as the primary data for measuring changes in student performance. Five-point behaviorally anchored rating scales were developed for each CRM skill area. Study-specific data collection forms were used by instructors to record observations of student skill levels in the flight immediately following crew practice in the PC-ATD for test group students. The same forms were used to document the CRM proficiency levels of students in the control group at the same point in flightline training.

- **CRM grades in student training records** for both groups were also included as metrics of effectiveness. First flight grades were extracted from these records as were sorties to the first fully proficient rating.

Events Impacting the Study

When the project began, the syllabus and training curriculum were built around an intact student crew (pilot, co-pilot, navigator, and flight engineer). Air Mobility Command changed the crew make-up at the beginning of 2005; the co-pilot initial qualification course was eliminated and flightline training events were decoupled across crew-specific courses. The test plan called for full student cockpit crews to go through the PC-ATD immediately prior to their first tactical low-level flight in the aircraft. Now four separate classes must arrive on the flight-line within a two-day period for that utilization strategy to work. It took five months before the first full student crew with this proper mix arrived on the flight-line at the same time. The left seat student pilot in this first crew found the simulator difficult to control given limited C-130 flying experience. The handling characteristics in the PC-ATD were assessed by flightline instructors and found to differ sufficiently from those of the aircraft to risk the potential for negative training for the pilot flying, especially during takeoff and landing. Finally, the FE station supported checklist participation, but few FE interfaces with aircraft systems were operational. This resulted in very low workloads for student FEs.

At that point it was decided to focus PC-ATD instruction on student navigators and right seat student pilots. Following this change, scheduling students for data PC-ATD training was more manageable. An instructor pilot flew the mission from the left seat and instructed the students from that position during the simulator session. The student pilot not-flying could then focus on CRM behaviors and interactions with the navigator.

Subjects

Eleven student right seat pilots (formerly copilots) and 11 student navigators provided the data reported in this study. All 22 were in the tactical phase of C-130 mission qualification training. None had previous C-130 or airdrop experience. Five student navigators and six right-seat student pilots received instruction in the PC-based CRM trainer. The in-flight performance of six other student navigators and five-right seat pilots who did not receive PC-ATD instruction (the control group) was rated for comparison purposes.

Given the small sample sizes in this study, it was prudent to assess the comparability of subjects prior to the airdrop phase of training. The Commander's Summary in each student folder was reviewed. This provides a compilation of all ATD performance grades prior to the airdrop phase of flightline training. Most student ATD grades were Good or Excellent but a few Conditional and Unsatisfactory grades were also reported. Eight percent of treatment group navigator grades and 13% of treatment group pilot grades were excellent in earlier ATD training. Similarly, 10% of control navigator and 10% of control group pilot grades were Excellent. Conditional grades were received by three control group students and four treatment group students. Two treatment group and no control group students had received Unsatisfactory grades prior to airdrop training. In summary, the treatment and control groups seemed to be very well matched. If there was any bias, it was minor and to the advantage of the control group.

The PC-ATD

Logan, Couvillion, and Clemons (2005) described the PC-ATD and the underlying engineering considerations in detail. An overview is provided here. Crew stations were provided for left-seat pilots, right-seat pilots, navigators, and flight engineers. An instructor operator station (IOS) was also part of the system. The training station for student pilots is shown in Figure 1. The FE would sit immediately behind the pilots. Not shown (but located to the right) would be the navigator station and the IOS. The original study objective was to investigate the ability of a low-cost, PC-based training device to improve teamwork skills. The system requirements documents specified sufficient fidelity to practice the cockpit crew coordination tasks associated with visual and instrument airdrop missions. The focus was not on training a C-130 crew to fly the aircraft, but rather on training inter-crew responsibilities throughout the mission. The training system needed to facilitate the interactions among the crew members that would be experienced in flight; it did not need to meet the standards of a Level C flight simulator nor fly identically to the actual aircraft.

Figure 1: The C-130 PC-Based CRM Trainer

In response to these stated requirements, the proposed system concept used Microsoft® Flight Simulator (MSFS) and supporting application programming interfaces to fulfill as many functional requirements as possible, including aircraft handling characteristics and out-the-window scene generation. COTS hardware components, including liquid crystal displays with touch screen interfaces, were used to support the crew interactions with the simulated flight instrumentation. While MSFS included C-130 flight characteristics, the airdrop environment involved several unusual parameters. For example, the low level airdrop scenario required the aircraft to maintain altitude at sub-100 knot airspeeds. This particular domain proved challenging throughout development. Several days were spent fine tuning aircraft performance in conjunction with Air Force subject matter experts (SMEs) to optimize the match between ATD and actual aircraft handling characteristics.

Another area that required high fidelity was the self contained navigation system (SCNS). Due to the complexity of the system, only a portion of the SCNS was replicated, targeting functionality in the NAV and TUNE pages. These two sections of the system allow the crew to set navigation aid frequencies and radios, and monitor route and scheduling information. Aspects of SCNS that were not relevant to the task were left out, leaving some menu options on the replicated SCNS non-functional. The physical structure for the PC-ATD was built using a modular aluminum framework system to keep the configuration flexible. The use of a modular system proved very beneficial for the proof-of-concept development as it allowed the structure to be designed "on the fly" to minimize construction and modification costs and material waste. The modular components also simplified disassembly for transport and the system could be easily modified or reconfigured.

For the out-the-window view, two 50" plasma displays were selected. The original design called for three 42" plasma displays, but due to the width of the screens, this would have put the edge of each display almost directly in front of the pilot and copilot. Air Force SMEs decided the extra horizontal field of view did not justify the visual obstruction that would have been unavoidable with three displays. A more important concern was the vertical field of view directly ahead of the aircraft for the low level navigation. In airdrop missions, it was important that visual

cues and landmarks on the ground are visible just ahead of the aircraft and the 50" plasma displays provided greater vertical coverage of these features than the original 42" screens would have provided.

Procedures - Training

A 4-hour time period was allotted for PC-ATD training including planning, briefing, a SKE airdrop, a visual airdrop, and debrief. The training session started with a description of what the simulator was designed to accomplish and what was expected of each crew position during the upcoming training session. It was stressed that the training was voluntary and any feedback would be greatly appreciated. The student crew was given all the necessary paperwork required to fly the mission.

Students were given a Form 280 that laid out the mission profile, communication frequencies, and formation call signs. The student pilot was given a low-level chart for the visual portion of the mission and a set of route drawings to fill out and give to the instructor pilot. The navigator was given a set of charts to fly the mission, route logs to plan the leg times, and the drop zone mosaics needed to plot their Computed Airdop Release Point (CARP). Students were given one hour to compile and prepare everything needed to conduct the briefing and fly the mission.

The PC-ATD mission started with the simulated aircraft on a SKE route in the number two wing position, and it continued through the SKE airdrop. Following the SKE airdrop, the crew finished the drop checklist and the SKE route was terminated. In the PC-ATD, students flew both SKE and visual routes to the All American drop zone. In the aircraft, students flew to the Blackjack drop zone, which is a more challenging drop zone due to more subtle terrain features and fewer cultural features in the area.

A 10-minute break allowed the instructor to reset the simulator for the visual low-level route. This route started from takeoff at Little Rock AFB. The crew was given threat information in code after takeoff, requiring them to exercise crew coordination skills in flight while passing navigation control from the navigator to the right-seat pilot so the navigator could to plot the threat and see if it was a factor given their route of flight. The navigator then took back navigation responsibility and the right-seat pilot plotted the threat. The two crewmembers compared their threat plots and looked for agreement. The navigator then updated the crew on the threat location and if it was a factor in their mission. The mission then proceeded to the airdrop. A PC-ATD design limitation required the use of the SCNS to support visual airdrops because the forward visibility in the simulator is such that items on the ground disappear two miles (or 40 seconds) prior to aircraft arriving at the point due to the configuration of the visual display.

Flightline training for C-130 student pilots occurs in two phases. The first phase focuses on basic C-130 operations and normally involves 3-5 flights. Phase II training addresses tactical airlift skills, including the targeted knowledge and skills associated with airdrop missions. The beginning of flightline training for navigators coincided with the beginning of Phase II flightline training for pilots. C-130 PC-ATD training occurred immediately prior to the first flight (an airdrop mission) for all student navigators and at the beginning of Phase II flightline training for right-seat student pilots. Navigator training and pilot Phase II training typically involves 6-12 flights, with airdrops occurring throughout this sequence.

Procedures—Data Collection

Participant feedback. Instructor feedback forms were developed in conjunction with 314 AW SMEs. They were administered to students following their first in-flight airdrop training mission. Student feedback forms were developed in a similar fashion, but were filled out by participating treatment group instructors once at the end of the data collection phase of the study.

Instructor ratings of CRM proficiency. Specialized CRM proficiency data collection forms were developed with inputs from 314 AW flightline instructors to capture student CRM skill levels for each crew position (pilot, copilot, navigator, flight engineer) during their first actual flight following instruction in the PC-ATD. Ultimately, only copilot and navigator data were collected in sufficient quantity to analyze results. Instructors rated inflight CRM proficiency both by CRM skill area and overall, and the forms generally followed the mission timeline for ease of use by instructors.

Aircrew Training Records. Electronic grade books contain a summary of each student's ground training, written evaluations, and hands-on performance evaluations. Instructors grade students' proficiency in each of the six CRM skill areas inflight by sortie. These grades are available in each student's flying training summary. Two grades were

observed. An S referred to satisfactory progress toward an eventual (P) proficiency rating. From these records, numbers of sorties to the first P grade were extracted for each student, for each of the six CRM skill areas.

RESULTS

Results are organized around the types of data that were collected. First, we describe the results of the surveys administered to instructors and students to address user reactions to the PC-ATD. Next, we present instructor ratings of CRM proficiency during the first actual in-flight airdrop based on our detailed CRM data collection forms. Results were used to assess the transfer of PC-ATD training to flight environment by comparing CRM ratings during the first actual airdrop for students who received PC-ATD instruction with CRM ratings of students who did not receive such training. Finally, we examine the CRM aspects of the aircrew training records to explore impacts on subsequent student performance.

User Reactions to the PC-ATD

Instructor Ratings. The first 15 items in the instructor survey were 5-point Likert scale questions that addressed issues of training effectiveness, fidelity/usability, and reliability. These items are summarized in left-hand column of Table 2. On the 5-point scale, a “5” corresponded to a very positive rating, and “1” a very negative rating. We computed the average (mean) ratings for each item across the 12 instructors who had used the PC-ATD; the results are shown in the right-hand column of Table 2. Importantly, the entire range of the scale was used, as respondents provided both “1’s” and “5’s” to some items with many of the “1” ratings referring to downtime from equipment malfunctions. Fifteen “5” ratings were distributed across effectiveness items plus adequacy of cues for instrument airdrops. There was considerable variability in the ratings across instructors. Two instructors provided overall average ratings of 3.0 and one instructor had an average of 4.0. The average rating, across all items and all instructors, was 3.5.

To gauge the degree to which the ratings were positive or negative, we compared the mean rating of each item across instructors against the scale mid-point of 3.0. A t-test of significance was used for each item, and a Bonferroni adjustment for multiple tests was applied (Harris, 1994). Items whose average ratings were significantly above the 3.0 midpoint can be considered to be rated positively based on this conservative test. As indicated by the asterisked items in Table 2, five aspects of the low-fidelity trainer received significantly positive ratings by the instructors.

Table 2. Instructor Ratings

Survey Item	Mean Rating
Effectiveness of Training Experience	
Overall CRM training	3.6
Task Management Skills	4.0*[1]
Communication Skills	3.6**[2]
Crew Coordination Skills	3.9*
SA Skills	3.2
Mission Planning/Briefing Skills	3.9*
Airdrop performance enhancement	3.3
Equipment operation skill	3.2
Effective use of time	3.8*
Fidelity/Usability of the Device	

[1] Ratings marked with an asterisk (*) were significantly higher ($p < .05$) than the scale midpoint of 3.0 based on a Bonferroni-adjusted t-test with 12 degrees of freedom.
[2] Ratings marked with a double asterisk (**) were marginally higher statistically ($.05 < p < .10$)

OTW Cues	3.3
Controls/Displays	3.3
Airdrop Cues	3.8*
Instructor Operating Station	3.5
Reliability of the Device	
PC-based trainer working	2.8
No downtime	2.6

These included its ability to: train task management skills, train crew coordination skills, train mission planning/briefing skills, provide effective use of the students and instructors' time, and provide usable airdrop cues with reasonable fidelity. A sixth item, the PC-ATD's ability to train communication skills, achieved a marginal level of statistical significance. For these six aspects, there were no "1" or "2" ratings from individual instructors, and "4" or "5" ratings were given by the majority of raters. Though none of the other survey items achieved significance, it is noteworthy that *none* of these items received a significantly negative mean rating. Indeed, only two items, those involving reliability, received an average rating below 3.0. The reasons behind these ratings can be gleaned through an examination of the comments, as discussed below.

Instructor Comments. The last few items on the survey asked additional, open-end questions concerning the effectiveness, fidelity, and reliability of the training device. While several instructors took advantage of these questions to explain some specific problems or areas of concern, the bulk of our qualitative investigation was based on face-to-face interviews that we conducted with three of the instructors toward the end of the project. Their opinions were quite insightful, and form the basis for the conclusions presented below.

First, and in line with the rating data, instructors consistently indicated that the PC-ATD demonstrated clear *potential* to enhance some CRM skills, most notably task management, crew coordination, planning/briefing, and communication. At a general level, a student's experience with the trainer showed them how "rushed they're going to be" inflight, so it gives them some familiarity with the brisk pace of events during airdrops. Navigators and co-pilots could practice the specific calls they make to one another, helping them hone cadence, vocabulary, and timing. Practice in making and listening to radio calls was also touted, as was gaining an understanding of the tactical sequence of events and basic switchology skills. While the visual fidelity of the device was somewhat limited, the prevailing view was that *any* opportunity to practice checklists and gain experience with team/crew interactions is going to yield positive dividends when students get into the aircraft. Indeed, instructors considered the trainer to be a very cost- and time-effective way to provide such training, *if* students are given enough repetitions to hone the necessary communication, checklist, and interaction skills.

On the other hand, in its present form, device reliability and ease of use were less than desired. Problems during system startup (e.g., the screens were often not lined up after startup) ultimately resulted in the loss of three crews from the study. Indeed, the frequent downtime and continual need to restart was largely responsible for the low ratings that were obtained in the areas of reliability and effectiveness. In addition, several of the Flight Management System pages (e.g., several of the airdrop pages) were not available in the trainer, limiting some of the tactics that could be conducted. This limitation may have been partly responsible for the lower ratings obtained in the area of situation awareness.

It was generally perceived that the FEs did not have enough to do during the simulation session, since there was really no system to operate, only checklists to be run. While even this skill, checklist operation, was of some value (See Table 1), the length of the training session with only this task to perform resulted in considerable boredom for most FE students.

Despite these problems and limitations, the majority instructors felt it was worth their time and that of their students to engage in training with the system as indicated by a "4" or "5" rating for this item. Consequently, they would recommend continued use of the system, although it was desirable to fix some of the aforementioned problems with reliability and ease of use before considering wider scale implementation.

Student Ratings. The first 17 items in the student feedback survey were 5-point Likert scale-type questions that addressed issues of training effectiveness, fidelity, usability, reliability, and planning. Like their counterparts in the instructor survey, a "5" corresponds to a very positive rating (strongly agree with) and a "1" a very negative rating

(strongly disagree with). The topics are listed in the left column of Table 3. Ten students completed the survey, all having served as subjects in the test (trainer) condition. Four were co-pilots, four were navigators, and two were flight engineers.

The mean rating for each survey item was computed across the 10 students. These are shown in Table 3. As was the case with instructors, the students used the entire scale range in their responses, with several students having recorded either a "1" or a "5." The few "1" ratings consistently reflected downtime due to equipment problems. There was notable variability in the ratings across students, as the average rating ranged from a low of 2.8-2.9 to a high of 4.1. The overall student average rating, 3.5, was identical to that obtained for the instructors. Because of low sample size, it was not feasible to compare ratings across crew positions. However, there was some evidence of slightly lower ratings for the FE. This would be consistent with the comment data, where it was evident (and confirmed by instructors) that there was simply less for the FEs to do in the PC-ATD compared to the other crew positions.

Once again, the absolute level of positiveness in the ratings was assessed by comparing each average to the scale midpoint of 3.0 using conservative Bonferroni adjustments. As indicated by the asterisked items in the table, students gave the following items ratings higher than the scale mid-point: crew coordination, pre-mission planning, debriefing, communication, task management, and ICDU and SKE control panel usage. The positive ratings for communication, task management and crew coordination are consistentwith the instructor ratings. In addition, students seemed to be enamored with the device's ability to promote mission planning, briefing, and debriefing skill development. Variability of ratings differed considerably different across items and as a result, some items with high ratings (training effectiveness of mission planning/briefing and effective use of student

Table 3. Student Ratings

Survey Item	**Mean Rating**
Effectiveness of Training Experience	
Enhanced my overall performance	3.5
Learning CRM aspects of airdrops	3.4
Crew Coordination Skills	4.1*
Communication Skills	3.6*
Situation Awareness Skills	3.5
Task Management Skills	3.7**
Mission Planning and Briefing	3.8
Radio procedures and radio operation	2.6
Effective use of time	3.7
Fidelity/Usability of the Device	
OTW Cues	3.1
Primary instrument displays – visual airdrops	3.1
Primary instrument displays – instrument airdrops	3.6
ICDU & SKE control panels	3.9**
Reliability of the Device	
PC-based trainer worked reliably	3.1

No downtime	2.9
Mission Preparation & Debriefing	
Planning & briefing was effective	4.2*
Debriefing was effective	4.2*

time) did not meet the criterion for statistically significant deviation from the mid point. Only two items (radio operations/procedures and downtime) received average ratings below 3.0. But again, these deviations below a “3” rating did not even come close to being statistically significant. Thus, for the most part, students—like their instructors—were fairly satisfied with the training opportunities provided by the device, and in some cases, were quite enthusiastic about them.

Student Performance Inflight

PC-ATD impacts on first flight CRM. Instructors provided detailed ratings of CRM proficiency of both treatment and control group students based on their performance during the first inflight airdrop flight following PC-ATD instruction. To assess the impact of instruction in the PC-ATD, our initial statistical analysis addressed the overall CRM ratings given to treatment group and control group students in their first in-flight airdrop mission. For this comparison, we de-coupled students from their crew assignments and treated each student as an individual. Because interpretation of the mean differences between the test and control groups requires that we make some inferences regarding the meaningfulness of partial scale values (e.g., a 3.3 average rating vs. 2.8 in this study), it was useful to consider other ways of expressing the effects found in the data. To that end, we performed a statistical analysis on the number of students who received an overall CRM score of “2” (marginal) or “1” (poor) on their first airdrop flight. We believed that this index had practical meaning, since instructors would be required to intervene more, demonstrate more aspects of crew coordination, and in general, spend more time on remedial instructing for students whose CRM proficiency is at a “2” or lower compared to students whose CRM proficiency meets (“3”) or exceeds (“4”) the standard level of proficiency.

Using overall ratings of CRM proficiency as our index, we computed numbers of control students who were rated as a “2” on the airdrop flight. For the control group, that number was 5 out of 11, or 45% (.45). Taking .45 as the proportion of students who would be expected to display a “2” or lower level of CRM proficiency, we then computed the corresponding number for the treatment group and assessed the probability of obtaining that proportion by chance, using a binomial distribution (Miller & Freund, 1965).

The number of students in the treatment group receiving an overall category CRM rating of “2” or lower was, in fact, 0, as all subjects received overall ratings of “3” or higher. The probability of observing this number, 0, by chance from an underlying binomial distribution with p=.45 and N=11 is .0014, which exceeded the alpha level of .05, thus indicating a statistically significant difference. This overall statistical treatment/control group difference gave us “permission” to examine the data in more detail to discern the loci of our effect (Harris, 1994).

First flight CRM proficiency by skill area. Figure 2 depicts the mean first flight CRM rating based on detailed CRM ratings from instructors for the two conditions (treatment and control) pertaining to each of the six CRM categories. Each average is based on a total of 11 subjects. Mean ratings tended to be concentrated in the 2.5-3.5 range, but there is consistent superiority of the treatment condition that is present in all six categories. The size of the difference varies somewhat, being largest for SA (3.3 vs. 2.5 for a difference of .8) and smallest for Mission Evaluation (3.2 vs. 3.0).

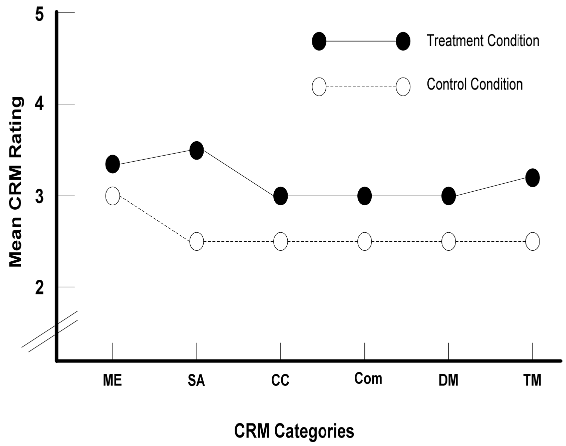

Figure 2. Mean In-Aircraft CRM Ratings for Treatment and Control Conditions by CRM Area.

To determine how these differences in CRM ratings stand up to statistical scrutiny, we first performed a multivariate analysis of variance (MANOVA) with two between-subject factors--crew position (navigator/co-pilot) and condition (treatment/control) and six dependent variables--the six CRM skill area ratings for each student. The MANOVA revealed that there was a significant effect of treatment condition ($F = 4.107$; df = 6, 13; $p < .016$). The effect of crew position was not significant ($F = 1.635$; df = 6, 18; $p < .215$) nor was the condition by crew position interaction ($F = .772$, df = 6, 13; $p < .606$).

Univariate follow-up tests revealed that this difference resided principally in two CRM categories, TM ($F = 9.257$; df = 1, 18; $p< .007$) and SA ($F = 6.890$; df = 1, 18; $p < .017$). However, marginal levels of significance were obtained in the targeted skill areas of communication ($F = 3.273$; df = 1, 18; $p <. 087$), and coordination ($F = 3.257$; df = 1, 18; $p < .088$), as well as in decision making ($F = 2.967$; df = 1, 18; $p < .102$).

Training Records--Flights to CRM Proficiency. Average flying sorties to the first proficient instructor grade in student records are shown in Figure 3. The large differences between mean sortie counts for navigators and pilots reflect two different course flows. Pilots have had 6-8 previous initial qualification sorties to develop basic C-130 flying skills that navigators did not receive. For both crew positions, the data in Figure 3 reflect numbers of airdrop sorties.

Sorties to first proficient rating in the two targeted skill areas (communication and coordination), were more than one sortie lower for student navigators who received training in the ATD (10.2 vs. 11.8 sorties). For pilots, average savings of .8 sortie were observed. Sorties to proficiency for task management and decision making/ risk assessment were also lower for students who received the four-hour ATD instruction. No difference was observed for mission planning for either crew position, but this may reflect a lack of emphasis on planning in this phase of inflight training rather than a lack of transfer. A sign test was performed on the pattern of mean differences in the CRM data in Figure 3. Specifically, we asked the statistical question of, out of 12 tests, what is the probability that 9 of the means would be smaller for the test group with 3 ties (Miller & Freund, 1965). Application of the sign test (Siegel, 1956) reveals that this outcome is highly unlikely ($p < .002$, $N = 9$). Thus, we have another piece of evidence supporting the existence of a persistent training effect associated with the low-fidelity CRM trainer.

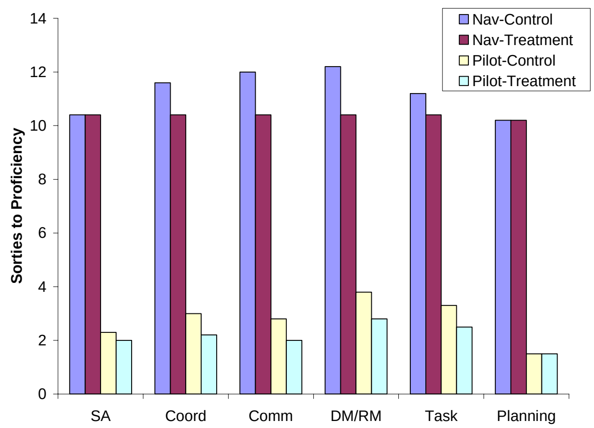

Figure 3: Mean sorties to first Proficient Rating from Flight Instructors for CRM Skills

CONCLUSIONS / RECOMMENDATIONS

Our original proof-of-concept study design was organized around Kirkpatrick's (1960) four-stage training effectiveness evaluation model: (1) assess trainee reaction to the training; (2) measure learning progress in the targeted training treatment; (3) evaluate performance in the intended work environment; and (4) measure training impact in terms of benefits to the organization. Training effectiveness studies often address the first two stages--the perceived value of the training and the degree to which the to-be-trained knowledge-skills-attitudes (KSAs) are actually learned in the device being evaluated. The availability of targeted KSAs on the job is less frequently addressed, and measuring benefits to the organization is rarely accomplished. This pattern is especially common with CRM training, which has been almost universally adopted throughout aviation despite surprisingly little empirical evidence linking this training to improved mission performance or safety (See, for example, Salas, et al., in press; Ilgen, 1999).

To assess the first stage (user perceptions of the PC-ATD), we surveyed both students and instructors, who both gave positive ratings to the PC-ATD for communication, crew coordination, and task management. On the other hand, instructor pilot reactions to the flight handling characteristics of the PC-ATD contributed to a decision to eliminate pilots flying from the study.

To address learning in the PC-ATD, the original test plan called for half the students in the treatment condition to accomplish a visual airdrop mission followed by a SKE airdrop mission in the PC-based trainer while the other half would accomplish a SKE airdrop mission followed by a visual airdrop mission. As external factors greatly reduced treatment group crew availability, the resulting smaller sample size no longer supported this differential treatment within treatment cells. As a result, we did not address the second stage of Kirkpartick's (1960) model. Fortunately, we were able to address both the third and fourth stages: transfer to the flight environment (initial inflight performance following ATD instruction) and benefits to the organization (sorties to proficiency). In both cases, the results were positive for the two crew positions studied.

The bottom line in this study was that the ability of the PC-based ATD (driven by MSFS) to support C-130 whole crew training was mixed. On the positive side, this PC-based technology appeared to provide a very effective environment in which team coordination skills can be trained for students not flying the aircraft (right-seat pilots and navigators). In fact, there appeared to be close to a one-to-one correspondence between a 4-hour training session in this PC-ATD and a decrease of one inflight sortie in terms of progressing toward proficiency.

For pilots flying, however, differences between MSFS handling characteristics and those of the aircraft itself were viewed by instructors as being sufficient to risk negative transfer of training for students who were still learning how to control a C-130, even after attempting to fine tune flight parameters and eliminating takeoffs and landings from airdrop training scenarios. In addition, recent changes in course flows at the C-130 schoolhouse resulted in crew positions entering the airdrop phase of training at varying times, substantially limiting the utility of any whole-crew trainer, including this ATD.

Training for pilots flying currently drives the generation of training flights at the 314th AW. As a result, return on investment depends upon the ability to reduce flying training needs for pilots flying. A PC-based, high-fidelity C-

130 handling package has been developed and is available. Assessing the ability of an ATD with this enhanced capability would be a logical next step if scheduling can be adjusted to have all crew positions entering the airdrop phase of training at the same time. Increasing interest in ways to reduce the need to fly the aging C-130E aircraft at Little Rock AFB may justify revisiting PC-based technology to shape effective teamwork behaviors, especially if enhanced flying characteristics are added.

Finally, a few words are in order pertaining to evaluating CRM training effectiveness. Salas and his colleagues (in press) strongly argued for the use of a multilevel approach to evaluating training outcomes (i.e., reactions, learning, behavior, and organizational impact). We agree. Each data source that we considered (user feedback, behaviorally anchored data collection forms, and student records provided both unique insights and the ability to corroborate findings across independent sources. Salas's final conclusion was "that more robust research, training and evaluations are needed so that we can fully grasp the impact that CRM is having in the community. At this point, we believe the tools to determine impacts are there; what we need are a mandate, access to data, and the resources to make it happen." Again, we concur. We were particularly encouraged to find student record data (grades and instructor comments) to be both valuable and efficient sources on information as part of our broader data collection effort and strongly recommend considering both in future training effectiveness studies.

REFERENCES

Air Force Instruction 11-290 (2001). *Cockpit/crew resource management training program.* Pentagon: HQ Air Force/XOOT.

Harris, R.J. (1994). *ANOVA: An analysis of variance primer*. Itasca, IL: F.E. Peacock Publishers, Inc.

Helmreich, R.L. & Foushee, H.C. (1993). Why crew resource management? Empirical and theoretical bases of human factors training in aviation. In E.L. Wiener, B.G. Kanki, and R.L. Helmreich, *Cockpit resource management*, San Diego, CA: Academic Press.

Hendy, K.C., Thompson, M.M., Fraser, W.D., Jamieson, D.W., Comeau, P. Mack, C.I., Paul, M.A., & Brooks, C.J. (1998). Human Factors of CC-130 operations Volume 1: Final report of the DCIEM/ATG study team. Defense and Civil Institute of Environmental Medicine, Toronto, Ontario.

Hendy, K.C. & Ho, G. (1998). Human Factors of CC-130 operations Volume 5: Human factors in decision making. Defense and Civil Institute of Environmental Medicine, Toronto, Ontario.

Ilgen, D.R. (1999). Teams embedded in organizations: Some implications. *American Psychologist*, *54*, 129-139.

Kirkpatrick, D. L. (1960). Techniques for evaluating training programs. *Journal of the American Society of Training Directors*, *14*, 13-18, 28-32.

Komich, J. (1997). CRM training: Which crossroads to take now? In R. S. Jensen & L. A. Rakovan (Eds.), *Proceedings of the 9th International Symposium on Aviation Psychology* (pp. 541-546). Columbus, OH: The Ohio State University.

Leonard, Y. (2005). *AETC study plan Project 03-03, C-130 desktop trainer*. Randolph AFB TX: Air Education and Training Command Studies and Analysis Squadron.

Logan, R.C., Couvillion, W.C., & Clemons, L. (2005). Exploring the development and effectiveness of a low cost "microsim" approach for teaching C-130 crew resource management skills. In *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference,* Orlando FL.

Miller, I. & Freund, J.E. (1965). *Probability and statistics for engineers.* Englewood Cliffs, NJ: Prentice-Hall.

Montimy, R.J. (2005). USAF aviation safety: FY 2004 in review. http://safety.kirtland.af.mil.

Nullmeyer, R., Spiker, V.A., Wilson, D. & Deen, G. (2003). Key crew resource management behaviors underlying C-130 aircrew performance. In *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference*, Orlando FL.

Nullmeyer, R., Stella, D., Flournoy, J. & White, D. (2003). Using Air Force mishap data to improve C-130 CRM training. In *Proceedings of the International Symposium on Aviation Psychology*, Dayton, OH.

Salas, E., Wilson, K.A., Burke, C.S., Wightman, D.C. & Howse, W. (In press). Crew resource management training and practice: A brief review, lessons learned and needs. Submitted to *Review of Human Factors and Ergonomics (*RHFE) series (Vol. 2).

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.

Simmon, D. A., (1997). How to fix CRM. In R. S. Jensen & L. A. Rakovan (Eds.), *Proceedings of the 9th International Symposium on Aviation Psychology* (pp. 550-553). Columbus, OH: The Ohio State University.

White, D.J. (2006). Improving CRM training in the U.S. Navy. Presented at the *Aerospace Medical Association Conference*, Orlando FL.